



Langevin-based Strategy for Efficient Proposal Adaptation in Population Monte Carlo

Víctor Elvira, Emilie Chouzenoux

► To cite this version:

Víctor Elvira, Emilie Chouzenoux. Langevin-based Strategy for Efficient Proposal Adaptation in Population Monte Carlo. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, May 2019, Brighton, United Kingdom. 10.1109/ICASSP.2019.8682284 . hal-02431677

HAL Id: hal-02431677

<https://hal.archives-ouvertes.fr/hal-02431677>

Submitted on 8 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LANGEVIN-BASED STRATEGY FOR EFFICIENT PROPOSAL ADAPTATION IN POPULATION MONTE CARLO

Víctor Elvira and Émilie Chouzenoux†*

* IMT Lille Douai & CRISAL (UMR CNRS 9189), Villeneuve d’Ascq (France).

† LIGM (UMR CNRS 8049), Univ. Paris Est (France).

CVN, INRIA Saclay, CentraleSupélec (France).

ABSTRACT

Population Monte Carlo (PMC) algorithms are a family of adaptive importance sampling (AIS) methods for approximating integrals in Bayesian inference. In this paper, we propose a novel PMC algorithm that combines recent advances in the AIS and the optimization literatures. In such a way, the proposal densities are adapted according to the past weighted samples via a local resampling that preserves the diversity, but we also exploit the geometry of the targeted distribution. A scaled Langevin strategy with Newton-based scaling metric is retained for this purpose, allowing to adapt jointly the means and the covariances of the proposals, without needing to tune any extra parameter. The performance of the proposed technique is clearly superior in two numerical examples at the cost of a reasonable computational complexity increment.

Index Terms— Importance sampling, Monte Carlo methods, population Monte Carlo, Langevin dynamics, Newton algorithm, stochastic optimization.

1. INTRODUCTION

In many problems of science and engineering, the goal is in inferring some unknown parameter from a set of data. In Bayesian inference, a posterior distribution of the unknown parameter conditioned to the data is built. Unfortunately, for most applications of interest, the posterior is intractable and must be approximated. Importance sampling (IS) is a popular family of Monte Carlo methods [1, 2]. Random samples are simulated from a proposal distribution and weighted accordingly so estimators of moments of the posterior distribution can be obtained. However it is a challenging problem to choose a good proposal in advance, and adaptive IS (AIS) procedures must be performed [3, 4].

AIS algorithms have received a particular attention in the recent years due to their ability to iteratively improve the proposal, and hence to increase the performance of IS. AIS can be applied to a large class of problems due to its flexibility

and weak requirements [4, 5]. The population Monte Carlo (PMC) algorithms are a family of AIS algorithms [6, 7, 8, 9] that implement a resampling procedure for adapting the location parameters proposal pdfs from previous weighted samples. However, the state-of-the-art PMC methods suffer from two known limitations. First, due to the resampling step, the path degeneracy endangers the diversity in the exploration of the space. The LR-PMC in [9] reduces this problem by resampling locally among disjoint sets of samples at the expense of worsening the local exploration w.r.t. a global resampling (see more details in [9]). Second, most of PMC methods only adapt the location parameter but are unable to fit the scale parameter.

Recent works in the literature have investigated the use of gradient steps in the proposal adaptation. For instance, [10] presents an AIS without resampling that exploits the gradient and the Hessian and introduces some artificial repulsion among proposals. In [11, 12], the proposal mean is adapted using one iteration of the unadjusted Langevin algorithm (ULA) [13] initialized in the samples values. In these methods, the covariance is also adapted, using autocorrelation of past samples [11], or second-order information [10, 12]. Langevin-based schemes have also been considered as acceleration strategies in MCMC algorithms [14, 15, 16]. The idea is to combine ULA with a Metropolis-Hastings step, giving rise to the so-called Metropolis adjusted Langevin algorithm (MALA). The advantage is that a sample arising from the proposal is more likely drawn from a highly probable region, which is beneficial to the acceptance rate. The performance of MALA can be improved by introducing in the drift term a scaling matrix depending on the current sample value, in order to adapt the proposal to the local structure of the target density [17, 18, 19]. Several strategies have been investigated for the construction of the scaling matrix in MALA, relying on second-order information [20, 18], Fisher metric [17] or majorization-minimization strategy [19].

In this work, we propose a new scaled Langevin PMC (SL-PMC) approach, where the proposals are adjusted using a ULA step with scaling metric adapted to the sample values resulting from a local resampling strategy. In contrast to the

V.E. and É.C. acknowledge support from the *Agence Nationale de la Recherche* of France under PISCES (ANR-17-CE40-0031-01) and MAJIC (ANR-17-CE40-0004-01) projects.

mentioned works, here the mean and covariance adaptation are performed jointly, with the advantage of fitting the proposal distributions locally, boosting the exploration and increasing the performance. A Newton strategy, associated to a backtracking linesearch is proposed for the construction of the Langevin drift term. We show on two sets of numerical examples that this novel methodology enhances the local adaptation without endangering the diversity of the proposals.

The rest of the paper is structured as follows. Section 2 introduces the problem setting, the AIS framework, and the Langevin diffusion. In Section 3, we present the proposed method. Finally, we show some numerical examples in Section 4 and conclude in Section 5.

2. MOTIVATING FRAMEWORK AND BACKGROUND

2.1. Bayesian inference

Let us consider a vector of unknowns $\mathbf{x} \in \mathbb{R}^{d_x}$ that is related to a set of data $\mathbf{y} \in \mathbb{R}^{d_y}$ through a statistical model represented by the likelihood function $\ell(\mathbf{y}|\mathbf{x})$. Prior information on \mathbf{x} is contained in $p_0(\mathbf{x})$. Then, the posterior distribution (also called target) is represented by

$$\tilde{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})p_0(\mathbf{x})}{Z(\mathbf{y})} \propto \pi(\mathbf{x}|\mathbf{y}), \quad (1)$$

where $\pi(\mathbf{x}|\mathbf{y}) = \ell(\mathbf{y}|\mathbf{x})p_0(\mathbf{x})$ is a non-negative function, and $Z(\mathbf{y}) = \int \pi(\mathbf{x}|\mathbf{y})d\mathbf{x}$.¹ In many applications, the goal consists in computing a moment w.r.t. the posterior distribution as in

$$I = \int h(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x} = \frac{1}{Z} \int h(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \quad (2)$$

where h is any integrable function w.r.t. $\tilde{\pi}(\mathbf{x})$. However, in many scenarios, Eq. (2) cannot be computed because either the integral is intractable and/or Z is not available.

2.2. Population Monte Carlo

Population Monte Carlo (PMC) is a family of adaptive importance sampling (AIS) algorithms for the approximation of integrals of the form of Eq. (2). In the following, we briefly describe the basic importance sampling (IS) method and a state-of-the-art PMC algorithm.

2.2.1. Importance sampling (IS)

In its basic version, IS simulates K samples as $\mathbf{x}_k \sim q(\mathbf{x})$, $k = 1, \dots, K$, where $q(\mathbf{x})$ is the *proposal* distribution. An importance weight is assigned to each sample as $w_k = \frac{\pi(\mathbf{x}_k)}{q(\mathbf{x}_k)}$, $k = 1, \dots, K$. Under mild conditions, the integral of Eq. (2)

¹We now drop \mathbf{y} in Z , $\pi(\mathbf{x})$, and $\tilde{\pi}(\mathbf{x})$ in order to alleviate the notation.

can be approximated by the (unbiased) unnormalized IS (UIS) estimator $\hat{I} = \frac{1}{KZ} \sum_{k=1}^K w_k h(\mathbf{x}_k)$, if Z is known. In the general case where Z is unknown, the (consistent) self-normalized IS (SNIS) estimator is implemented

$$\tilde{I} = \sum_{k=1}^K \bar{w}_k h(\mathbf{x}_k). \quad (3)$$

It is well known that the optimal proposal for the UIS estimator is $q(\mathbf{x}) \propto |h(\mathbf{x})|\pi(\mathbf{x})$ [1, 2]. However, since it is usually impossible to know in advance the best proposal, adaptive mechanisms are employed.

2.2.2. Deterministic mixture PMC (DM-PMC)

The PMC family of algorithms is a popular implementation of the adaptive IS (AIS) method. The framework was proposed in [6], and arguably, its most important feature is the use of resampling procedures for adapting the proposal pdfs. The DM-PMC [9] is a recent PMC algorithm with high performance that will serve as basis for our description here and also for the numerical examples. In DM-PMC, the parameters of N proposal densities are initialized. In particular, the initial location parameters $\{\boldsymbol{\mu}_n^{(1)}\}_{n=1}^N$ will be adapted, while the parameters $\{\mathbf{C}_n\}_{n=1}^N$ are static (e.g., for Gaussian proposals, the location parameters are the means and the static parameters are the covariance matrices). At each iteration t , exactly K samples are simulated from each proposal $q_n^{(t)}(\mathbf{x})$. The NK samples receive an IS weight, and the weights are normalized. Finally, the N location parameters for the next iteration, $\{\boldsymbol{\mu}_n^{(t+1)}\}_{n=1}^N$, are simulated from the set of NK samples according to a resampling procedure. The DM-PMC is an advanced algorithm that outperforms the standard PMC in most scenarios due to more stable estimators (see [21]) and a resampling step that better preserves the diversity. The challenge of this step is in selecting the N next location parameters of the proposals from a pool of NK samples. Three strategies have been proposed for the generic DM-PMC [9, 22] with the names of global (GR-PMC), local (LR-PMC), and partial (PR-PMC) resampling PMC, depending on the type of resampling they implement. See more information in [22, Fig. 1].

2.3. Langevin diffusion scheme

A Q -dimensional Langevin diffusion [14] is a continuous-time Markov process $(\mathbf{x}(t))_{t \geq 0} \in \mathbb{R}^{d_x}$ defined through the following stochastic differential equation:

$$(\forall t \geq 0) \quad d\mathbf{x}(t) = \mathbf{b}(\mathbf{x}(t))dt + \mathbf{A}^{1/2}(\mathbf{x}(t))d\mathbf{B}(t), \quad (4)$$

with $(\mathbf{B}(t))_{t \geq 0} \in \mathbb{R}^{d_x}$ a Brownian motion, and for every $\mathbf{x} \in \mathbb{R}^{d_x}$, $\mathbf{A}(\mathbf{x}) \in \mathbb{R}^{d_x \times d_x}$ a symmetric definite positive (SDP) scaling matrix. The drift term $\mathbf{b}(\mathbf{x}) = (b_i(\mathbf{x}))_{i=1}^{d_x}$ is related

to the target density (assumed to be differentiable) and to the scaling matrix through

$$b_i(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^{d_x} A_{i,j}(\mathbf{x}) \frac{\partial \log \pi(\mathbf{x})}{\partial x_j} + |\mathbf{A}(\mathbf{x})|^{\frac{1}{2}} \sum_{j=1}^Q \frac{\partial}{\partial x_j} \left(A_{i,j}(\mathbf{x}) |\mathbf{A}(\mathbf{x})|^{-\frac{1}{2}} \right), \quad (5)$$

with $1 \leq i \leq d_x$. The Langevin diffusion process has stationary distribution $\tilde{\pi}$. Using the Euler's discretization scheme with stepsize θ yields ULA [13, 23], defined, for every $t \in \mathbb{N}$ as

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \theta \mathbf{b}(\mathbf{x}^{(t)}) + \sqrt{\theta} \mathbf{A}^{1/2}(\mathbf{x}^{(t)}) \boldsymbol{\omega}^{(t+1)}, \quad (6)$$

where $(\boldsymbol{\omega}^{(t)})_{t \in \mathbb{N}} \in \mathbb{R}^{d_x}$ are i.i.d. realizations of zero-mean Gaussian noise.

3. SCALED LANGEVIN POPULATION MONTE CARLO

3.1. Proposed algorithm

We are now ready to present our new SL-PMC algorithm in Table 1. We initialize the N proposals with location parameters $\{\boldsymbol{\mu}_n^{(1)}\}_{n=1}^N$, scale parameters $\{\boldsymbol{\Sigma}_n^{(1)}\}_{n=1}^N$, and other static parameters $\{\boldsymbol{\nu}_n\}_{n=1}^N$. For T iterations, we proceed as follows. In step2(a), K samples are simulated from each proposal. An IS weight is associated to each sample in step 2(b) (note that we use the deterministic mixture weight with whole mixture of proposals in the denominator [21]). In step 2(c), the local resampling of [9] is implemented in order to select the set of parameters $\{\tilde{\boldsymbol{\mu}}_n^{(t+1)}\}_{n=1}^N$. In particular, the n -th parameter $\tilde{\boldsymbol{\mu}}_n^{(t+1)}$ is resampled from the set of K samples generated by the proposal located at $\boldsymbol{\mu}_n^{(t)}$, i.e., from the set $\mathcal{X}_n^{(t)} = \{\mathbf{x}_{n,1}^{(t)}, \dots, \mathbf{x}_{n,K}^{(t)}\}$ with associated probabilities $\bar{w}_{n,k}^{(t)} = \frac{w_{n,k}^{(t)}}{\sum_{\ell=1}^K w_{n,\ell}^{(t)}}$, $k = 1, \dots, K$. Note that exactly one sample from each proposal survives from iteration t to $t+1$, which preserves the diversity and the local exploration of next step. In step 2(d), the mean of the proposal density at next iteration is updated using

$$\boldsymbol{\mu}_n^{(t+1)} = \tilde{\boldsymbol{\mu}}_n^{(t+1)} + \frac{1}{2} \theta_n^{(t+1)} \mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)}) \nabla \log \pi(\tilde{\boldsymbol{\mu}}_n^{(t+1)}), \quad (7)$$

where $\mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)})$ is an SDP matrix of $\mathbb{R}^{d_x \times d_x}$. The scaled gradient term in (7) can be understood as a discretization of (5) under the assumption of a locally constant curvature [19]. Like in the ULA scheme (6), the covariance matrix of the proposal density is adapted according to

$$\boldsymbol{\Sigma}_n^{(t+1)} = \theta_n^{(t+1)} \mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)}). \quad (8)$$

After T iterations, we deliver the NKT weighted samples that can be used to build the estimators of Section 2.2.1.

Table 1. SL-PMC algorithm.

1. **[Initialization]:** Set $\sigma > 0$, $(N, K, T) \in \mathbb{N}^+$, $\{\boldsymbol{\nu}_n\}_{n=1}^N$. For $n = 1, \dots, N$, select the initial adaptive parameters $\boldsymbol{\mu}_n^{(1)} \in \mathbb{R}^{d_x}$ and $\boldsymbol{\Sigma}_n^{(1)} = \sigma^2 \mathbf{I}_{d_x}$.
2. **[For $t = 1$ to T]:**
 - (a) Draw K samples from each proposal pdf,
$$\mathbf{x}_{n,k}^{(t)} \sim q_n^{(t)}(\mathbf{x}; \boldsymbol{\mu}_n^{(t)}, \boldsymbol{\Sigma}_n^{(t)}, \boldsymbol{\nu}_n) \quad (9)$$
with $n = 1, \dots, N$, and $k = 1, \dots, K$.
 - (b) Compute the importance weights,
$$w_{n,k}^{(t)} = \frac{\pi(\mathbf{x}_{n,k}^{(t)})}{\frac{1}{N} \sum_{i=1}^N q_i^{(t)}(\mathbf{x}_{n,k}^{(t)})}. \quad (10)$$
 - (c) Resample N location parameters $\{\tilde{\boldsymbol{\mu}}_n^{(t+1)}\}_{n=1}^N$ from the set of NK weighted samples of iteration t using the local resampling strategy.
 - (d) Adapt the proposal parameters $\{(\boldsymbol{\mu}_n^{(t+1)}, \boldsymbol{\Sigma}_n^{(t+1)})\}_{n=1}^N$ according to (7)-(8).
3. **[Output, $t = T$]:** Return the pairs $\{\mathbf{x}_{n,k}^{(t)}, w_{n,k}^{(t)}\}$, for $n = 1, \dots, N$, $k = 1, \dots, K$ and $t = 1, \dots, T$.

3.2. Parameters setting

The parameters $(\theta, \mathbf{A}(\cdot))$ play an important role in our method, as θ determines the length of the proposed jumps, whereas the scale matrix $\mathbf{A}(\cdot)$ controls their direction. In the following, we propose a simple and efficient strategy for the setting of these values.

3.2.1. Scaling matrix.

In order to better infer the local curvature of the target, we propose to choose the scaling metric in our Langevin-based scheme relying on the information of the Hessian of $\log \pi$:

$$\mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)}) = \left(-\nabla^2 \log \pi(\tilde{\boldsymbol{\mu}}_n^{(t+1)}) \right)^{-1}. \quad (11)$$

Note that, despite the complexity increase resulting from the inversion step in (11), such scaling strategy has been observed to yield very satisfactory performance in the context of MALA [20, 18]. It has also been employed with success for the covariance adaptation in PMC methods [12, 10]. A difficulty is that for general densities $\tilde{\pi}$, the log-concavity assumption is not necessarily fulfilled and numerical issues may arise in the inversion of the Hessian matrix. We propose to overcome this issue by using $\boldsymbol{\mu}_n^{(t+1)} = \tilde{\boldsymbol{\mu}}_n^{(t+1)}$ and $\boldsymbol{\Sigma}_n^{(t+1)} = \sigma^2 \mathbf{I}_{d_x}$, for all $n \in \{1, \dots, N\}$ when the matrix (11) is not well-defined. Let us emphasize that, in our practical tests, the value retained for σ does not seem to have a critical role in the performance of our method (in contrast

to standard DM-PMC where this parameter choice may be sensitive).

3.2.2. Stepsize tuning.

If matrix $\mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)})$ is well defined, there still remains to set the associated stepsize value $\theta_n^{(t+1)}$. We propose to resort to a simple backtracking scheme. Starting with unit stepsize value, we reduce the stepsize by factor $\tau = 1/2$ until the condition below is met:

$$\pi(\tilde{\boldsymbol{\mu}}_n^{(t+1)} + \theta_n^{(t+1)} \mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)}) \nabla \log \pi(\tilde{\boldsymbol{\mu}}_n^{(t+1)})) \geq \pi(\tilde{\boldsymbol{\mu}}_n^{(t+1)}).$$

Note that this method was observed in our experiments to lead to better stability and exploration performance than the decreasing stepsize rule retained in [11].

4. NUMERICAL RESULTS

4.1. Mixture of Gaussians

Let us consider the bivariate multimodal target pdf given by $\tilde{\pi}(\mathbf{x}) = \frac{1}{5} \sum_{i=1}^5 \mathcal{N}(\mathbf{x}; \gamma_i, \mathbf{C}_i)$ for $\mathbf{x} \in \mathbb{R}^2$, with means $\gamma_1 = [-10, -10]^\top$, $\gamma_2 = [0, 16]^\top$, $\gamma_3 = [13, 8]^\top$, $\gamma_4 = [-9, 7]^\top$, $\gamma_5 = [14, -4]^\top$, and covariance matrices $\mathbf{C}_1 = [5, 2; 2, 5]$, $\mathbf{C}_2 = [2, -1.3; -1.3, 2]$, $\mathbf{C}_3 = [2, 0.8; 0.8, 2]$, $\mathbf{C}_4 = [3, 1.2; 1.2, 0.5]$ and $\mathbf{C}_5 = [0.2, -0.1; -0.1, 0.2]$. The goal is the approximation of the target mean $\mathbb{E}_{\tilde{\pi}}[\mathbf{X}] = [2.4, 3.4]^\top$, the second moment $\mathbb{E}_{\tilde{\pi}}[\mathbf{X}^2] = [101.04, 98.94]^\top$, and the normalizing constant $Z = 1$. We compare the performance of the novel SL-PMC algorithm with GR-PMC and LR-PMC, two variants of the DM-PMC algorithm [9]. For GR-PMC and LR-PMC, we set proposal covariances $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I}_{d_x}$, with $\sigma \in \{1, 3, 5\}$, while we take $\sigma = 5$ in SL-PMC. All methods are run with $N = 50$ proposals (randomly initialized in the square $[-4, 4] \times [-4, 4]$), $T = 20$ iterations, and $K = 20$ samples per proposal and iteration. Table 2 shows the relative mean square error (RMSE) of the AIS estimators, averaging the results of the second half of iterations. The novel SL-PMC outperforms the best setup of the DM-PMC by roughly an order of magnitude in the estimation of the three quantities.

	GR-PMC			LR-PMC			SL-PMC
	$\sigma = 1$	$\sigma = 3$	$\sigma = 5$	$\sigma = 1$	$\sigma = 3$	$\sigma = 5$	
Z	0.6419	42.1047	0.0289	0.2807	0.1309	0.1522	0.0014
$\mathbb{E}_{\tilde{\pi}}[\mathbf{X}]$	41.3552	8.0010	0.3583	5.4810	1.6225	0.4860	0.0238
$\mathbb{E}_{\tilde{\pi}}[\mathbf{X}^2]$	12.0858	10.0200	0.5253	6.5815	2.1486	0.6844	0.0556

Table 2. Example 4.1. Relative MSE in the estimation of Z , $\mathbb{E}_{\tilde{\pi}}[\mathbf{X}]$, and $\mathbb{E}_{\tilde{\pi}}[\mathbf{X}^2]$.

4.2. High-dimensional banana-shaped distribution

Let us consider the random variable (r.v.) $\mathbf{X} \in \mathbb{R}^{d_x}$, $d_x \geq 2$. This r.v. is a transformation from d_x -dimensional multivariate Gaussian $\bar{\mathbf{X}} \sim \mathcal{N}(\mathbf{x}; \mathbf{0}_{d_x}, \mathbf{C})$ with $\mathbf{C} = \text{diag}(c^2, 1, \dots, 1)$.

The transformed r.v. is computed as $X_j = \bar{X}_j$, $j \in \{1, \dots, d_x\} \setminus 2$, and $X_2 = \bar{X}_2 - b(\bar{X}_1^2 - c^2)$, with $c = 1$ and $b = 3$. This transformation leads to a complicated banana-shaped distribution with uncorrelated components [24, 25]. Figure 1 shows the MSE in the estimation of $\mathbb{E}_{\tilde{\pi}}[\mathbf{X}]$ for different dimensions $d_x \in \{2, 5, 10, 15, 20, 30, 40, 50\}$. In all methods, we set $N = 50$, $K = 20$ and $T = 20$. For both variants GR-PMC and LR-PMC, we display the performance with proposal covariances $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I}_{d_x}$, with $\sigma \in \{1, 3\}$ (the performance with $\sigma = 5$ is much worse and is not displayed). In general, the performance of all methods worsens when the dimension grows. Note that the SL-PMC outperforms the DM-PMC in all dimensions, keeping a substantial gap in terms of MSE.

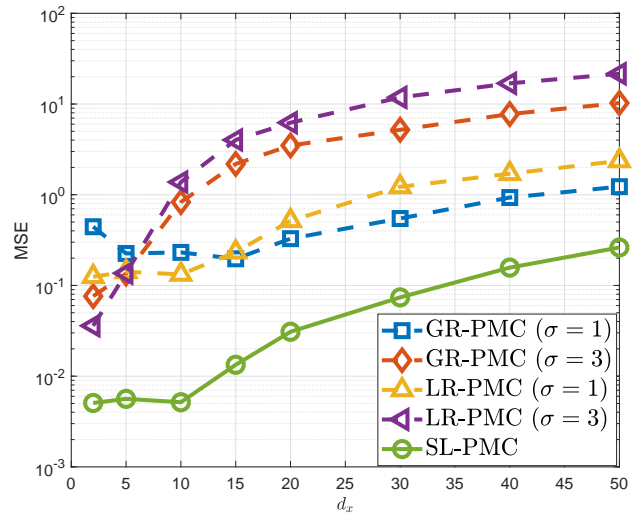


Fig. 1. Example 4.2. MSE in the estimation of $\mathbb{E}_{\tilde{\pi}}[\mathbf{X}]$ of the banana-shaped distribution versus the dimension d_x .

5. CONCLUSIONS

Population Monte Carlo algorithms are flexible adaptive importance sampling techniques for approximating distributions. Current state-of-the-art algorithms do not exploit enough the geometry of the posterior. In this paper, we have proposed a Langevin-based strategy with position dependent scaling metric, allowing to adapt jointly the means and the covariances of the proposals by accounting for second-order information about the target. This novel SL-PMC algorithm extracts the features of competitive PMC methods, adding new features coming from the optimization literature. Our method shows superior performance, especially in high-dimensional distributions, without the need of any extra parameter setting.

6. REFERENCES

- [1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.
- [2] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2004.
- [3] C. Andrieu and J. Thoms, “A tutorial on adaptive MCMC,” *Stat. Comput.*, vol. 18, no. 4, pp. 343–373, 2008.
- [4] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Míguez, and P. M. Djuric, “Adaptive importance sampling: The past, the present, and the future,” *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 60–79, 2017.
- [5] L. Martino, V. Elvira, D. Luengo, and F. Louzada, “Adaptive Population Importance Samplers: A General Perspective,” in *Proc. IEEE Sensor Array and Multichannel Sig. Process. (SAM 2016)*, Rio de Janeiro, Brazil, 10-13 July 2016.
- [6] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, “Population Monte Carlo,” *Journal of Comp. and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.
- [7] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert, “Adaptive importance sampling in general mixture classes,” *Stat. Comput.*, vol. 18, pp. 447–459, 2008.
- [8] P. Del Moral, A. Doucet, and A. Jasra, “Sequential Monte Carlo samplers,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 3, pp. 411–436, 2006.
- [9] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, “Improving Population Monte Carlo: Alternative weighting and resampling schemes,” *Sig. Process.*, vol. 131, no. 12, pp. 77–91, 2017.
- [10] V. Elvira, L. Martino, L. Luengo, and J. Corander, “A gradient adaptive population importance sampler,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2015)*, Brisbane, Australia, 19-24 April 2015, pp. 4075–4079.
- [11] I. Schuster, “Gradient importance sampling,” Tech. Rep., 2015, <https://arxiv.org/abs/1507.05781>.
- [12] M. Fasiolo, F. E. de Melo, and S. Maskell, “Langevin incremental mixture importance sampling,” *Stat. Comput.*, vol. 28, no. 3, pp. 549–561, 2018.
- [13] G. O. Roberts and L. R. Tweedie, “Exponential convergence of Langevin distributions and their discrete approximations,” *Bernoulli*, vol. 2, no. 4, pp. 341–363, Dec. 1996.
- [14] G. O. Roberts and O. Stramer, “Langevin diffusions and Metropolis-Hastings algorithms,” *Methodol. Comput. Appl. Probab.*, vol. 4, no. 4, pp. 337–357, 2002.
- [15] A. Durmus, E. Moulines, and M. Pereyra, “Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau,” *SIAM J. Imaging Sci.*, vol. 11, no. 1, pp. 473–506, 2018.
- [16] A. Schreck, G. Fort, S. Le Corff, and E. Moulines, “A shrinkage-thresholding Metropolis adjusted Langevin algorithm for Bayesian variable selection,” *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 2, pp. 366–375, March 2016.
- [17] M. Girolami and B. Calderhead, “Riemann manifold Langevin and Hamiltonian Monte Carlo methods,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 73, no. 91, pp. 123–214, Mar. 2011.
- [18] Y. Zhang and C. A. Sutton, “Quasi-Newton methods for Markov chain Monte Carlo,” in *Proc. Neural Information Processing Systems (NIPS 2011)*, Granada, Spain, 12-17 Dec. 2011, number 24, pp. 2393–2401.
- [19] Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, and J.C. Pesquet, “Majorize-Minimize adapted Metropolis-Hastings algorithm,” Tech. Rep., 2018, <https://hal.archives-ouvertes.fr/hal-01909153>.
- [20] J. Martin, C. L. Wilcox, C. Burstedde, and O. Ghattas, “A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion,” *SIAM J. Sci. Comput.*, vol. 34, no. 3, pp. 1460–1487, Jan. 2012.
- [21] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, “Generalized multiple importance sampling,” *Stat. Sci.*, 2018, <https://arxiv.org/abs/1511.03095>.
- [22] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, “Population monte carlo schemes with reduced path degeneracy,” in *Proc. IEEE Int. Work. Comput. Adv. Multi-Sensor Adap. Process. (CAMSAP 2017)*, 2017, pp. 1–5.
- [23] A. Durmus and E. Moulines, “Non-asymptotic convergence analysis for the unadjusted Langevin algorithm,” *Ann. Appl. Probab.*, vol. 27, no. 3, pp. 1551–1587, 2017.
- [24] H. Haario, E. Saksman, and J. Tamminen, “Adaptive proposal distribution for random walk metropolis algorithm,” *Comput. Stat.*, vol. 14, no. 3, pp. 375–396, 1999.
- [25] H. Haario, E. Saksman, and J. Tamminen, “An adaptive Metropolis algorithm,” *Bernoulli*, vol. 7, no. 2, pp. 223–242, April 2001.